

<https://helda.helsinki.fi>

Impact of the reference list features on the number of citations

Mammola, Stefano

2021-01

Mammola , S , Fontaneto , D , Martinez , A & Chichorro , F 2021 , ' Impact of the reference list features on the number of citations ' , Scientometrics , vol. 126 , no. 1 , pp. 785-799 . <https://doi.org/10.1007/s11192-020-03759-0>

<http://hdl.handle.net/10138/326141>

<https://doi.org/10.1007/s11192-020-03759-0>

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.



Impact of the reference list features on the number of citations

Stefano Mammola^{1,2} · Diego Fontaneto¹ · Alejandro Martínez¹ · Filipe Chichorro²

Received: 30 July 2020 / Accepted: 9 October 2020 / Published online: 29 October 2020
© The Author(s) 2020

Abstract

Many believe that the quality of a scientific publication is as good as the science it cites. However, quantifications of how features of reference lists affect citations remain sparse. We examined seven numerical characteristics of reference lists of 50,878 research articles published in 17 ecological journals between 1997 and 2017. Over this period, significant changes occurred in reference lists' features. On average, more recent papers have longer reference lists and cite more high Impact Factor papers and fewer non-journal publications. We also show that highly cited articles across the ecological literature have longer reference lists, cite more recent and impactful references, and include more self-citations. Conversely, the proportion of 'classic' papers and non-journal publications cited, as well as the temporal span of the reference list, have no significant influence on articles' citations. From this analysis, we distill a recipe for crafting impactful reference lists, at least in ecology.

Keywords Bibliography · Bibliometrics · Citations · Classic paper · Impact factor · Reference list · Self-citations

Introduction

As young scientists moving our first steps in the world of academic publishing, we were instructed by our mentors and supervisors on the articles to read and cite in our publications. "Avoid self-citations"; "Include as many papers published in *Nature* and *Science* as possible"; "Don't forget the classics"; and "Be timely! Cite recent papers" are all examples of such advice found in textbooks and blogs about scientific writing. Although these recommendations seem reasonable, they remain subjective as long as the effect of reference list features on citation counts of papers remains unknown.

The success of a scientific publication varies owing to a range of factors, often acting synergistically in driving its impact (Tahamtan et al. 2016). Apart from the scientific

✉ Stefano Mammola
stefano.mammola@helsinki.fi; stefano.mammola@cnr.it

¹ Molecular Ecology Group (MEG), IRSA – Water Research Institute, National Research Council, Verbania, Pallanza, Italy

² Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History (LUOMUS), University of Helsinki, Helsinki, Finland

content of the article itself, which ideally should be the only predictor of its impact, factors that correlate to the number of citations that an article accumulates over time include its accessibility (Gargouri et al. 2010; Lawrence 2001), the stylistic characteristics of its title (Bowman and Kinnan 2018; Fox and Burns 2015; Letchford et al. 2015; Murphy et al. 2019) and abstract (Freeling et al. 2019; Letchford et al. 2016; Martínez and Mammola 2020), the number of authors (Fox et al. 2016) and the diversity of their affiliations (Sanfilippo et al. 2018), and its availability as a preprint (Fu and Hughey 2019). Furthermore, the quality of a scientific publication should be related to the quality of the science it cites, but quantitative evidence for this remains sparse given that most analyses to date focused on a limited number of features of reference lists (e.g., Ahlgren et al. 2018; Didegah and Thelwall 2013; Fowler and Aksnes 2007; Fox et al. 2016; Haslam et al. 2008; Peters and van Raan 1994; Stewart 1983; Yu et al. 2014).

From a theoretical point of view, a reference list of high quality should be a balanced and comprehensive selection of up-to-date references, capable of providing a snapshot of the intellectual ancestry supporting the novel findings presented in a given article (May 1967). Scientists usually achieve this by conducting a systematic retrospective search to select all papers with content that is strongly related to that of the article, to be read, and potentially cited if deemed relevant. However, additional factors can drive the process of citing publications. For example, according to the constructivist theory on citations, one of the main reasons for picking up specific references over others is persuading the scientific community about the claims made in the paper by citing well-known authors in a field or articles published in top-tier journals (Tahamtan and Bornmann 2018).

An interesting recent attempt to evaluate the properties of a journal article reference list was made by Evans (2008). Using a database of > 30 million journal articles from 1945 to 2006 Evans showed that, over time, there has been a general trend to referencing more recent articles, channelling citations toward fewer journals and articles, and shortening the length of the reference list. Evans predicted that this way of citing papers “[...] may accelerate consensus and narrow the range of findings and ideas built upon”, an observation that generated subsequent debate (Gingras et al. 2009; Larivière et al. 2009; Von Bartheld et al. 2009). For example, in a heated reply to Evan’s report, Von Bartheld et al. (2009) argued that this claim was speculative because “[...] citation indices do not distinguish the purposes of citations”. In their view, one should consider the ultimate purpose of each citation and the motivation of authors when they decided which papers to cite.

Yet, it is challenging to disentangle all factors driving an author choice of citing one or another reference (Amancio et al. 2012; Tahamtan and Bornmann 2018; Wright and Armstrong 2011), especially when dealing with large bibliometric databases such as the one used by Evans (2008) to draw his conclusions. Probably the most recent and thorough attempt to explore this topic is the analysis by Ahlgren et al. (2018) focused on the effect of four reference list features on citations across nearly one million articles. They found that articles citing more references, more recent references, and more publications referenced in Web of Science accumulate, on average, more citations (Ahlgren et al. 2018).

Building upon this legacy, we set to examine the role of an even ampler set of reference list features in driving a paper citation impact. We extracted from Web of Science (Clarivate Analytics) all primary research journal articles published in low- to high-rank international journals in ecology in the last 20 years and generated unique descriptors of their reference lists (Table 1). We restricted our analysis to articles published in international journals in “Ecology” because, by focusing on a single discipline, it was possible to minimize the number of confounding factors. Moreover, this choice allowed us to incorporate

Table 1 Proxy variables used to characterize the reference list of the papers

Variable	Description	Construction
Length	Total number of reference items cited in the paper reference list	Sum of reference items cited
Non-journal items	Number of non-journal items cited in the reference list, such as books, theses, websites, and grey literature	Number of non-journal items cited divided by length
Self-citations	Proportion of self-citations in the reference list	Number of self-citations divided by length. Note that only first author self-citations are counted, namely those in which any of the authors of the paper appear as first author in items of the reference list
Temporal span	Temporal span of the reference list	Year of the most recent reference cited item minus year of the oldest cited item
Immediacy	Proportion of recent reference items cited	Number of papers published in the previous three years divided by length
Seminality	Proportion of seminal ecological papers cited	Number of cited items in the list “100 articles every ecologist should read” (Courchamp and Bradshaw 2018) divided by length
Total Impact Factor	Total Impact Factor of the reference list	Total Impact Factor is calculated using the journal Impact Factor of each reference at the year of publication, based on annual Journal Citation Reports by Web Of Science. To calculate the proportion, total Impact Factor is divided by the number of reference with an Impact Factor (length minus the number of non-journal items)

a unique descriptor of the reference list based on an analysis published in 2018 on seminal papers in ecology (Courchamp and Bradshaw 2018) (see “Seminality index” in Table 1).

We structured this research under two working hypotheses. First, if the quality of a scientific paper is connected to the reference it cites, we predict that, on average, articles characterized by a good reference list should accumulate more citations over time, where the goodness of a reference list is approximated via a combination of different indexes (Table 1). Second, we hypothesize that thanks to modern searching tools such as large online databases, bibliographic portals, and hyperlinks, the behavior through which scientists craft their reference lists should have changed in the Internet era (Gingras et al. 2009; Landhuis 2016). Thus, we predict that this change should be reflected by variations through time in the features of articles’ reference lists.

Methods

Criteria for articles inclusion

We extracted from Web of Science all primary research articles published in the ecological journals between 1997 and 2017 (Table 2). We excluded review and opinion papers, methodological papers, corrections, and editorials as these may have atypical reference lists. We downloaded articles on 19 August 2019, in Helsinki, using the browser Google Chrome (see Pozsgai et al. 2020). The year 1997 was chosen because approximately around this date the use of Impact Factor started to grow exponentially (Archambault

Table 2 Journal selected for the analysis

Journal name	Initial year	Temporal span selected	Total number of articles	Number of primary research articles
Acta Oecologica	1983	1997–2017	1571	1408
American Naturalist	1867	1997–2017	3417	2852
Austral Ecology	2000	2000–2017	1659	1434
Ecography	1978	1997–2017	2051	1743
Ecological Applications	1991	1997–2017	3641	3051
Ecology	1920	1997–2017	6584	5505
Ecology Letters	1998	1998–2017	2636	2098
Functional Ecology	1987	1997–2017	2889	2326
Global Change Biology	1995	1997–2017	4573	3937
Global Ecology and Biogeography	1993	1997–2017	1570	1377
Journal of Animal Ecology	1932	1997–2017	2639	2250
Journal of Applied Ecology	1964	1997–2017	2993	2407
Journal of Biogeography	1974	1997–2017	3541	2852
Journal of Ecology	1913	1997–2017	2603	2170
Molecular Ecology	1992	1997–2017	7853	6209
Oecologia	1968	1997–2017	6417	5446
Oikos	1949	1997–2017	4687	3812

and Larivière 2009). We selected only those ecological journals covering more than 16 years of the 21-years period considered, to be able to explore temporal trends with confidence. For example, *Nature Ecology and Evolution* (2016–ongoing) was excluded as it covered only 5 years of this temporal interval.

We generated seven descriptors of the features of reference lists and used these as variables in subsequent analyses. A description of each variable and the rationale for its construction are in Table 1. Note that most of the reference list features are expressed as proportions, to normalize variables to the number of papers cited in the reference list (Waltman and van Eck 2019).

Relationship between citations and reference list features

We conducted all analyses in R (R Core Team 2020). To test our first working hypothesis, we conducted regression-type analyses (Zuur and Ieno 2016). We initially explored our dataset following a standard protocol for data exploration (Zuur et al. 2009), whereby we: (1) checked for outliers in the dependent and independent variables; (2) explored the homogeneity of variables distribution; and (3) explored collinearity among covariates based on pairwise Pearson correlations, setting the threshold for collinearity at $|r| > 0.7$ (Dormann et al. 2013).

As a result of data exploration, we removed three outliers from articles citations corresponding to three papers cited over 6500 times in Web of Science, to remove their disproportionate effect on the analyses. We also observed that over 40% of the articles in our dataset were never cited, but since these represent “true zeros” (Blasco-Moreno et al. 2019) we didn’t apply zero-inflated models to infer citation patterns over time (Zuur et al. 2012). No collinearity was detected among the seven explanatory variables—all $|r| < 0.7$.

We used a Poisson generalized additive model to predict the expected pattern of citations over article age and expressed the number of citations as the Pearson residuals from the curve (Fig. 1h). To test which reference list features correlate with residuals in number of citations, we generated a linear mixed effect model by including journal identity and publication year as random terms to account for data non-independence. In other words, the choice of a mixed effect model design allowed us to take into account the confounding effect of the year of publication (temporal dependence) and of the journal in the estimation of regression coefficients (Silk et al. 2020). We fitted the linear mixed effect model with the R package “nlme” (Pinheiro et al. 2019), and validated models using residuals and fitted values (Zuur and Ieno 2016). To facilitate model convergence, we homogenized the distribution of our explanatory variables by log-transforming reference list size and temporal span, and square-root arcsin transforming all proportional variables (Crawley 2007).

Change in reference list features over time

We used linear mixed effect models to predict annual variations in reference list characteristics over time, including journal identity as a random factor. Seven models were constructed, one for each variable described in Table 1. In this case, as the seven variables were included as dependent variables, we didn’t log- and square-root arcsin transform them.

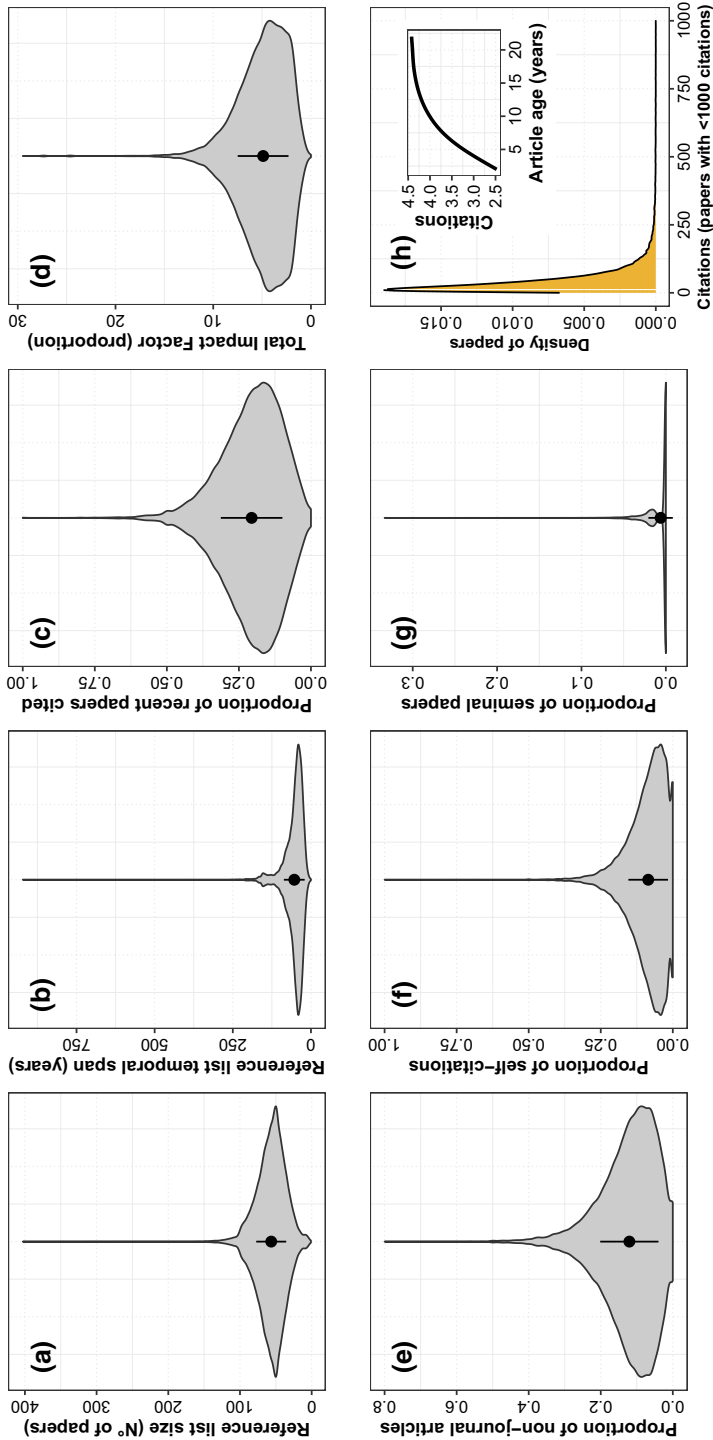


Fig. 1 Main numerical features of the reference list of ecological journals. **a–g** Violin plots showing the distribution of the seven numerical features of reference lists considered in this study. For each graph, the black dot and vertical bar are mean \pm standard deviation. **h** Distribution of citations among the articles considered in this study. Inset shows the predicted relationship between citations and articles age, based on the prediction of a generalized additive model

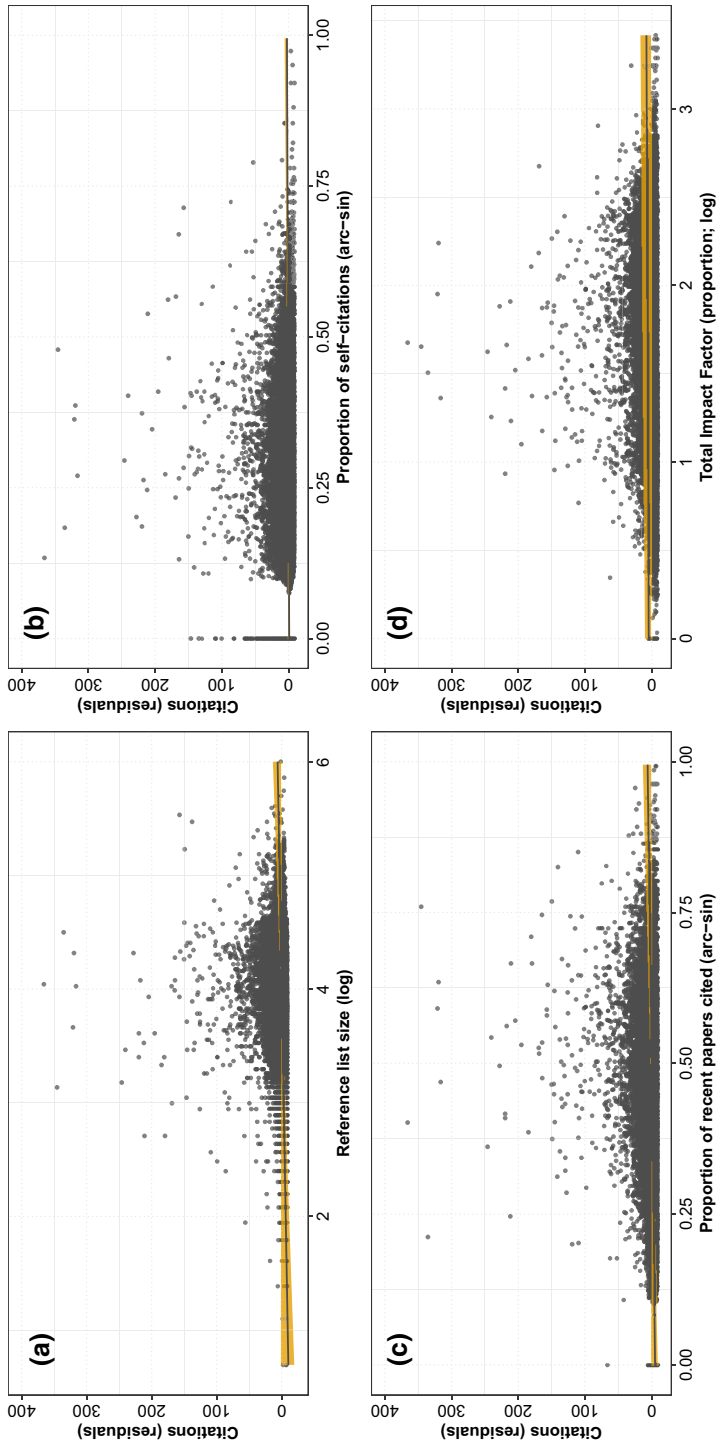


Fig. 2 Relationships between articles citation and reference list numerical features. Predicted relationships (filled lines) and 95% confidence intervals (orange surfaces) between the residuals of citation over articles' age and **a** length of the reference list, **b** proportion of self-citation, **c** proportion of recent papers cited, and **d** total Impact Factor of the reference list, according to the linear mixed effect models analysis. Variables are transformed to homogenize their distribution. Only fixed effects are shown

Fig. 3 Variations in reference list numerical features between 1997 and 2017. **a–g** Violin plots showing the annual variations in the seven numerical features of reference lists. For each graph, the black dot and vertical bar are mean \pm standard deviation. Insets show the predicted relationships (filled line) and 95% confidence intervals (orange surfaces) based on linear mixed models. Larger graph (**a–c**) illustrates non-flat temporal trends. Only fixed effects are shown

Results

Reference list features in ecology

After excluding non-primary research articles and omitting incomplete Web of Science records, we ended up with 50,878 unique papers distributed across the 17 journals that covered the temporal span from 1997 to 2017. The median size of the reference list in ecological journals is 54 cited items (range = 1–403) (Fig. 1a). Cited references cover a median temporal span of 45 years (0–922) (Fig. 1b). The mean proportion of recent papers in the reference lists is 0.21 (0–1); the proportion of non-journal items is 0.12 (0–0.8), whereas the average Impact Factor of the papers cited in reference lists is 4.9 (0–29.5) (Fig. 1). The mean proportion of self-citations is 0.07 (Fig. 1f) and the proportion of cited seminal papers is 0.006 (0–0.33) (Fig. 1g).

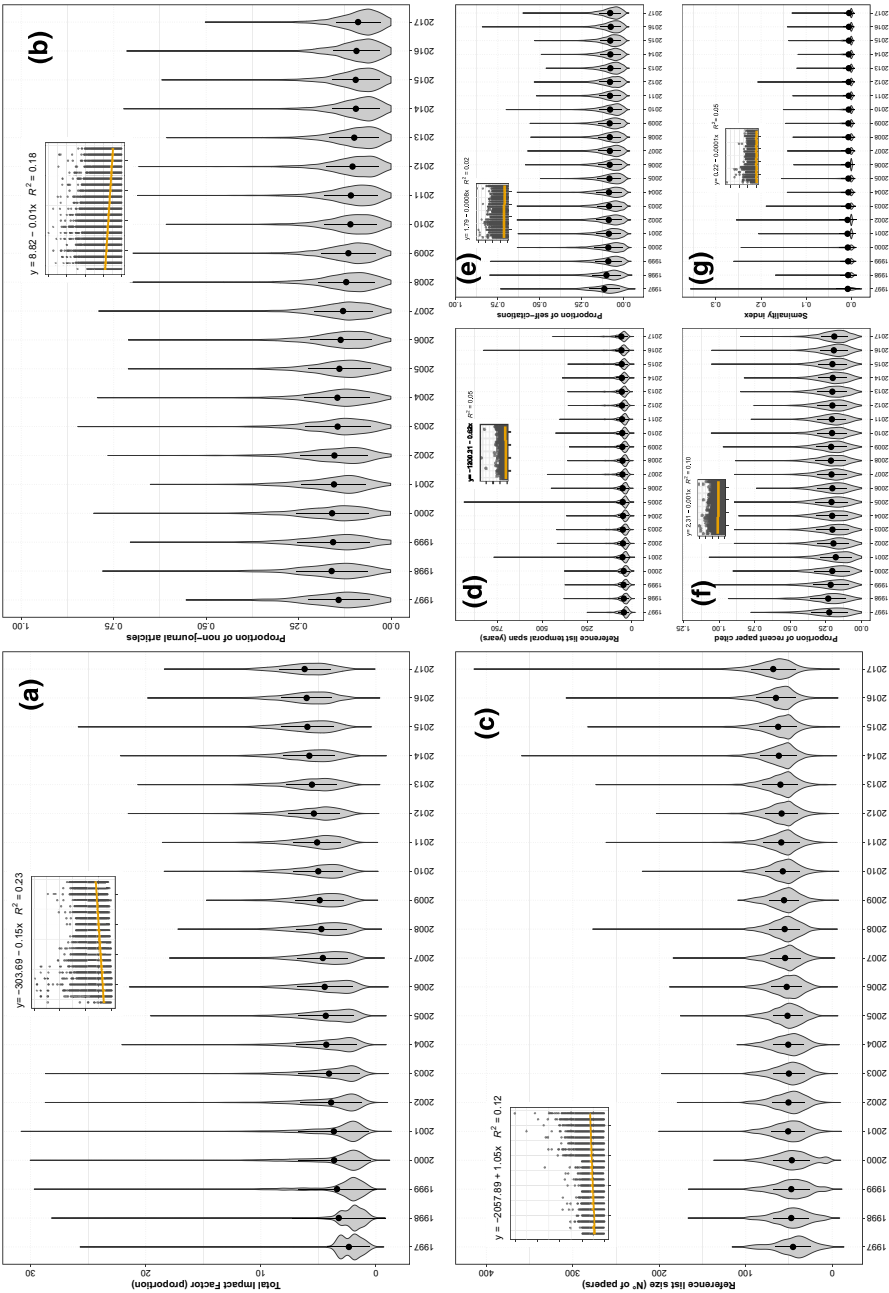
We predicted the expected curve of citations over article age with a Poisson generalized additive model. We observed a significant parabolic trend in the number of citations over time ($F=2724.8$; $p<0.001$), with number of citations reaching a plateau of ~ 4 after 10 years from publication (Fig. 1h, inset).

Relationship between reference list features and article citations

We observed a positive and significant relationship between citations of a paper (expressed as Pearson residuals) and the number of cited references (Estimated $\beta \pm \text{s.e.}$ 3.11 ± 0.12 , $t = 25.18$, $p < 0.001$), with articles with longer reference lists accumulating more citations over time (Fig. 2a). The number of citations also significantly increased with an increase in the proportion of self-citations (Estimated $\beta \pm \text{s.e.}$ 3.45 ± 0.34 , $t = 10.01$, $p < 0.001$; Fig. 2b) and reference list total Impact Factor (Estimated $\beta \pm \text{s.e.}$ 0.99 ± 0.12 , $t = 8.38$, $p < 0.001$; Fig. 2d). Furthermore, we found a positive relationship between citations and immediacy of the reference list; namely, articles citing a greater proportion of recent papers accumulated more citations over time (Estimated $\beta \pm \text{s.e.}$ 11.28 ± 0.39 , $t = 29.02$, $p < 0.001$; Fig. 2c). Proportion of non-journal items referenced, total temporal span of the reference list, and proportion of cited seminal papers had no significant effect on citations (non-journal items: Estimated $\beta \pm \text{s.e.}$ -0.22 ± 0.39 , $t = -0.54$, $p = 0.554$; Temporal span: -0.13 ± 0.35 , $t = -1.33$, $p = 0.164$; Seminality: 0.46 ± 0.644 , $t = 0.75$, $p = 0.470$).

Temporal variations in reference list features

Over the considered 21 years (1997–2017), the total Impact Factor of the reference list steadily and significantly increased. The average ($\pm \text{s.d.}$) Impact Factor of articles cited in the reference list was 2.35 ± 1.83 in 1997, and 6.19 ± 2.23 in 2017 (Fig. 3a). Yet, it is worth noting that the overall Impact Factor of scientific journals also significantly increased over the last 21 years, a feature that may have inflated this trend (Archambault and Larivière



2009). In parallel, the proportion of non-journal items referenced significantly decreased over time. In 1997, on average, non-journal items accounted for 14% of the reference list, while this value dropped to 8% in 2017 (Fig. 3b). We also observed that the number of cited items in the reference list steadily increased from an average of 45.3 ± 20.5 in 1997 to 68.2 ± 25.5 in 2017 (Fig. 3c). We observed stabler trends for the temporal span of the reference list (Fig. 3d), the proportion of self-citations (Fig. 3e), recent papers (Fig. 3f), and seminal papers (Fig. 3g). Models estimated parameters are in Fig. 3.

Discussion

Relationship between reference list features and citations

We showed that, on average, papers with longer reference lists are more cited across the ecological literature than papers with shorter reference lists, a result that parallels findings of previous studies (Ahlgren et al. 2018; Evans 2008; Fox et al. 2016; Onodera and Yoshikane 2015). One explanation is that longer reference lists may make papers more visible in online searches. Also, articles with longer reference lists may address a greater diversity of ideas and topics (Fox et al. 2016), thus containing more citable information. Furthermore, a long reference list may attract *tit-for-tat* citations, that is, the tendency of cited authors to cite the papers that cited them (Webster et al. 2009). It is interesting to emphasize that this result directly questions the practice of most journals to set a maximum in the number of citable references per manuscript. Since most journals are switching to online-only publishing systems where space limitation is not an issue, this limitation seems unjustified.

Consistently with previous evidence (Ahlgren et al. 2018), we also found that papers citing a greater proportion of recent articles are, on average, more cited. In parallel, we found that referencing articles with a high Impact Factor has a further positive effect on citations. Citing recent references published in top-tier journals generally implies that scientists are working on ‘hot’, timely eco-evolutionary topics. The latter frequently end up published in journals with greater Impact Factor, which on average attract a greater share of citations.

Furthermore, we found that papers including a greater proportion of self-citations are more highly cited—although it must be noted that there was a saturating point around 0.6–0.7 proportion of self-citations, after which papers were poorly cited. In general, given that excessive self-citations are usually despised and discouraged, the result of our analysis may come at a surprise. However, previous evidence on this issue is controversial (Aksnes 2003; Lievers and Pilkey 2012). For example, a study on the scientific production of Norway between 1981 and 1996 found the least cited papers accounted for the highest share of self-citation (Aksnes 2003). A following analysis, still based on Norwegian scientific productivity, showed that the more one cites him or herself, the more one is cited by other scholars (Fowler and Aksnes 2007).

It is true that self-citations are sometimes unjustified, used by authors as a way to increase their scientific visibility and to boost their citation metrics (Fowler and Aksnes 2007). An irrelevant self-citation breaking the flow of a paragraph (e.g., Mammola 2020) is an instructive example of this behavior. However, self-citations are an integrant part of scientific progress, as they usually reflect the cumulative nature of individual research (Ioannidis 2015; Mishra et al. 2018). As Glänzel et al. (2004) pointed out “... the absolute lack of self-citations over a longer period is just as pathological as an always-overwhelming share”. Indeed, 88% of the papers in our dataset included at least one self-citation. This

may ultimately lead to accumulating more citations because papers that are part of a bigger research line are often more visible and citable.

According to our analyses, other features of the reference list have not significant effect on citations. Probably, the least intuitive result is a lack of relationship between the number of cited seminal papers and the number of citations. We generated the list of seminal papers using a recent expert-based opinion paper, providing a list of the 100 “must-read” articles in ecology (Courchamp and Bradshaw 2018). A manuscript citing any of those classical papers should focus, on average, on broader and long-debated topics in ecology, and therefore it is expected to receive more citations. But this is not the case. If one assumes that the number of citations for a paper is an index of its importance for the field, such a result may question the “must-read” value of some of the articles included in Courchamp and Bradshaw (2018) compilation. However, most of these seminal papers are relatively old and they have inspired more recent studies, which may be cited instead of the original ones. As an alternative—or complementary—explanation, we noted that only 23% of articles in our database cited one or more seminal papers, which may partly affect the estimation of statistical significance.

Change in reference list features over time

We observed significant changes in the features of articles’ reference lists from 1997 to 2017. We argue here that most of these changes are directly related to a shift in the academic publishing behaviors of the Internet era (Fire and Guestrin 2019) from browsing paper in print to searching online through the use of hyperlinks (Evans 2008; Gingras et al. 2009). While the volume of available scientific information has grown exponentially (Landhuis 2016), retrieving relevant bibliography has become simpler and quicker thanks to online searching tools (Gingras et al. 2009). For instance, some studies reported that, as a result of the growing availability of papers online, the citation rate of older papers is actually rising in recent years (Krampen 2010; Wu et al. 2012). This enhanced accessibility of literature seemingly explains why, on average, the length of reference lists across ecological journals has steadily increased.

The last two decades have also seen an exponential rise in the use of journal metrics, especially the Impact Factor (Archambault and Larivière 2009), and the consequent desire of authors to publish in high-ranking journals and cite papers published therein. This may explain why we observed a significant increase in the total Impact Factor of reference lists over time. Concomitantly, there has been a reduction in the number of non-journal items cited in reference lists. In general, both these features are a direct product of the changes in academic publishing behaviors of the “publish or perish” era. More and more authors are now exploring new ways to maximize the impact of their publications (Doubleday and Connell 2017; França and Monserrat 2019; Freeling et al. 2019). Citing papers with higher Impact Factors and a lower proportion of non-journal items may be perceived as an effective way to achieve such a goal.

Caveats of the study

There are two main caveats that need to be made when interpreting the results. First, it is well-understood that a range of non-scientific factors beyond reference list features can affect citation counts (Tahamtan et al. 2016). One may argue that there could be synergistic

effects between the characteristics of the reference list and other features of a paper. For example, the number of co-authors in a paper could feedback to affect reference list features, insofar as each author may be more familiar with a different corpus of literature and propose citing different papers as a result. At the same time, the size of a reference list is often directly proportional to the length of the main text. Similar interactions remain unexplored.

A second caveat is that we here focused only on ecological literature. The degree to which our results are transferable to other disciplines could be explored using a database covering multiple disciplines.

Concluding remarks

While we are writing, identifying and citing the most relevant articles that provide the scientific foundation for our research questions is not trivial. Time is against us: most researchers are overloaded by academic duties and have busy schedules, preventing them to read classic papers and to keep up with the latest advances in the main and nearby fields of research. Memory failures, perhaps increased by the haste of finishing the manuscript on time, do not help either. Accordingly, reference lists are almost inevitably characterized by faulty citations, including incorrect references, quotation errors, and omitted relevant papers (Taylor 2002; Wright and Armstrong 2011). In a more cynical reasoning, May (1967) even argued that omissions of relevant papers might be due to the simple fact that “[...] the author selects citations to serve his scientific, political, and personal goals and not to describe his intellectual ancestry.”

But once we accept that making the perfect reference list is not possible, three heuristic rules will help us get close to it:

- (1) Size matters. Not only in terms of reference list but also in the number of characters (Ball 2008; Elgendi 2019). Investing extra resources into reading others’ research improves the scientific basis of the study while building argumentation links with relevant manuscripts, making the paper more visible and useful to peers.
- (2) Hotness. During the last 20 years, we have seen the advent of the Internet and changes in the way information is found, read, and spread. Keep track of impactful latest research, even exploiting novel tools such as social media (Thelwall et al. 2013) and blogs (Saunders et al. 2017), is a crucial premise to produce highly citable science (Ahlgren et al. 2018).
- (3) Narcisism. Not only self-citations directly increase the citations of past work, but they have been shown to improve the chances of being cited by others (Fowler and Aksnes 2007). Furthermore, the probability of self-citation increases with professional maturity in a given field of study, showing that that is a direct consequence of the cumulative nature of individual research (Mishra et al. 2018). Judicious use of self-citation may thus be beneficial in terms of citation impact.

Acknowledgements Thanks to Pedro Cardoso for a useful discussion. We are grateful to two anonymous referees for their constructive comments and for suggesting several additional references that enlarged our reference list. In line with our results, this hopefully will affect citation counts of this paper. SM acknowledge support from the European Commission through Horizon 2020 research and innovation programme (Grant No. 678193). FC was supported by the Finnish Kone Foundation.

Author contributions SM conceived the study. SM and FC designed methodology. FC mined data from Web of Science. SM and FC developed code for data processing. SM performed the analyses, with suggestions by FC, DF, and AM. SM wrote the first draft. All authors contributed to the writing of the manuscript through comments and additions.

Funding Open access funding provided by University of Helsinki including Helsinki University Central Hospital.

Data availability All data used to generate this study can be freely downloaded from Web Of Science. The cleaned database and R script to generate the analysis are available in Figshare (<https://doi.org/10.6084/m9.figshare.12941639.v1>).

Compliance with ethical standards

Conflict of interest The authors declare no competing financial interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahlgren, P., Colliander, C., & Sjögarde, P. (2018). Exploring the relation between referencing practices and citation impact: a large-scale study based on Web of Science data. *Journal of the Association for Information Science and Technology*, 69(5), 728–743. <https://doi.org/10.1002/asi.23986>.
- Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics*, 52, 235–246. <https://doi.org/10.1023/A:1021919228368>.
- Amancio, D. R., Nunes, M. G. V., Oliveira, O. N., & da Costa, L. F. (2012). Using complex networks concepts to assess approaches for citations in scientific papers. *Scientometrics*, 91, 827–842. <https://doi.org/10.1007/s11192-012-0630-z>.
- Archambault, É., & Larivière, V. (2009). History of the journal Impact factor: Contingencies and consequences. *Scientometrics*, 79(3), 635–649. <https://doi.org/10.1007/s11192-007-2036-x>.
- Ball, P. (2008). A longer paper gathers more citations. *Nature*, 455(7211), 274. <https://doi.org/10.1038/455274a>.
- Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M., & Castells, E. (2019). What does a zero mean? Understanding false, random and structural zeros in ecology. *Methods in Ecology and Evolution*, 10(7), 949–959. <https://doi.org/10.1111/2041-210X.13185>.
- Bowman, D., & Kinnan, S. (2018). Creating effective titles for your scientific publications. *VideoGIE*, 39(3), 260–261. <https://doi.org/10.1016/j.vgie.2018.07.009>.
- Courchamp, F., & Bradshaw, C. J. A. (2018). 100 articles every ecologist should read. *Nature Ecology & Evolution*, 2(2), 395–401. <https://doi.org/10.1038/s41559-017-0370-9>.
- Crawley, M. J. (2007). *The R book*. Hoboken: Wiley. <https://doi.org/10.1002/9780470515075>.
- Didegah, F., & Thelwall, M. (2013). Determinants of research citation impact in nanoscience and nanotechnology. *Journal of the American Society for Information Science and Technology*, 64(5), 1055–1064. <https://doi.org/10.1002/asi.22806>.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Doubleday, Z. A., & Connell, S. D. (2017). Publishing with objective charisma: breaking science's paradox. *Trends in Ecology and Evolution*, 32(11), 803–805. <https://doi.org/10.1016/j.tree.2017.06.011>.
- Elgendy, M. (2019). Characteristics of a highly cited article: a machine learning perspective. *IEEE Access*, 7, 87977–87986. <https://doi.org/10.1109/ACCESS.2019.2925965>.

- Evans, J. A. (2008). Electronic publication and the narrowing of science and scholarship. *Science*, 321(5887), 395. <https://doi.org/10.1126/science.1150473>.
- Fire, M., & Guestrin, C. (2019). Over-optimization of academic publishing metrics: observing Goodhart's Law in action. *GigaScience*, 8(6), giz053. <https://doi.org/10.1093/gigascience/giz053>.
- Fowler, J. H., & Aksnes, D. W. (2007). Does self-citation pay? *Scientometrics*, 72(3), 427–437. <https://doi.org/10.1007/s11192-007-1777-2>.
- Fox, C. W., & Burns, C. S. (2015). The relationship between manuscript title structure and success: editorial decisions and citation performance for an ecological journal. *Ecology and Evolution*, 5(10), 1970–1980. <https://doi.org/10.1002/ece3.1480>.
- Fox, C. W., Paine, C. E. T., & Sauterey, B. (2016). Citations increase with manuscript length, author number, and references cited in ecology journals. *Ecology and Evolution*, 6(21), 7717–7726. <https://doi.org/10.1002/ece3.2505>.
- França, T. F. A., & Monserrat, J. M. (2019). Writing papers to be memorable, even when they are not really read. *BioEssays*, 2019, 1900035. <https://doi.org/10.1002/bies.201900035>.
- Freeling, B., Doubleday, Z. A., & Connell, S. D. (2019). How can we boost the impact of publications? Try better writing. *Proceedings of the National Academy of Sciences*, 116(2), 341–343. <https://doi.org/10.1073/pnas.1819937116>.
- Fu, D. Y., & Hughey, J. J. (2019). Releasing a preprint is associated with more attention and citations for the peer reviewed article. *eLife*, 8, 52646. <https://doi.org/10.7554/eLife.52646>.
- Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., et al. (2010). Self-Selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE*, 5(10), e13636. <https://doi.org/10.1371/journal.pone.0013636>.
- Gingras, Y., Larivière, V., & Archambault, É. (2009). Literature citations in the internet era. *Science*, 323(5910), 36. <https://doi.org/10.1126/science.323.5910.36a>.
- Glänzel, W., Bart, T., & Balázs, S. (2004). A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics*, 59(1), 63–77. <https://doi.org/10.1023/B:SCIE.0000013299.38210.74>.
- Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., et al. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, 76(1), 169–185. <https://doi.org/10.1007/s11192-007-1892-8>.
- Ioannidis, J. P. A. (2015). A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation. *Journal of Psychosomatic Research*, 78(1), 7–11. <https://doi.org/10.1016/j.jpsychores.2014.11.008>.
- Krampen, G. (2010). Acceleration of citing behavior after the millennium? Exemplary bibliometric reference analyses for psychology journals. *Scientometrics*, 83(2), 507–513. <https://doi.org/10.1007/s11192-009-0093-z>.
- Landhuis, E. (2016). Scientific literature: information overload. *Nature*, 535(7612), 457–458.
- Larivière, V., Gingras, Y., & Archambault, É. (2009). The decline in the concentration of citations, 1900–2007. *Journal of the American Society for Information Science and Technology*, 60(4), 858–862. <https://doi.org/10.1002/asi.21011>.
- Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 411(6837), 521. <https://doi.org/10.1038/35079151>.
- Letchford, A., Moat, H. S., & Preis, T. (2015). The advantage of short paper titles. *Royal Society Open Science*. <https://doi.org/10.1098/rsos.150266>.
- Letchford, A., Preis, T., & Moat, H. S. (2016). The advantage of simple paper abstracts. *Journal of Informetrics*, 10(1), 1–8. <https://doi.org/10.1016/J.JOI.2015.11.001>.
- Lievers, W. B., & Pilkey, A. K. (2012). Characterizing the frequency of repeated citations: The effects of journal, subject area, and self-citation. *Information Processing & Management*, 48(6), 1116–1123. <https://doi.org/10.1016/j.ipm.2012.01.009>.
- Mammola, S. (2020). On deepest caves, extreme habitats, and ecological superlatives. *Trends in Ecology & Evolution*, 35(6), 469–472. <https://doi.org/10.1016/j.tree.2020.02.011>.
- Martínez, A., & Mammola, S. (2020). Specialized terminology limits the reach of new scientific knowledge. *bioRxiv*, 2020.08.20.258996. <https://doi.org/10.1101/2020.08.20.258996>.
- May, K. O. (1967). Abuses of citation indexing. *Science*, 156(3777), 890–892. <https://doi.org/10.1126/science.156.3777.890-a>.
- Mishra, S., Fegley, B. D., Diesner, J., & Torvik, V. I. (2018). Self-citation is the hallmark of productive authors, of any gender. *PLoS ONE*, 13(9), e0195773. <https://doi.org/10.1371/journal.pone.0195773>.
- Murphy, S. M., Vidal, M. C., Hallagan, C. J., Broder, E. D., Barnes, E. E., Horna Lowell, E. S., et al. (2019). Does this title bug (Hemiptera) you? How to write a title that increases your citations. *Ecological Entomology*, 44(5), 593–600. <https://doi.org/10.1111/een.12740>.

- Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology*, 66(4), 739–764. <https://doi.org/10.1002/asi.23209>.
- Peters, H. P. F., & van Raan, A. F. J. (1994). On determinants of citation scores: A case study in chemical engineering. *Journal of the American Society for Information Science*, 45(1), 39–49. [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1%3c39::AID-ASIS5%3e3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1%3c39::AID-ASIS5%3e3.0.CO;2-Q).
- Pinheiro, J., Bates, D., & DebRoy S, S. (2019). Linear and nonlinear mixed effects models. *R Package Version 3.1-140*. <https://cran.r-project.org/package=nlme>
- Pozsgai, G., Lövei, G. L., Vasseur, L., Gurr, G., Batáry, P., Korponai, J., et al. (2020). A comparative analysis reveals irreproducibility in searches of scientific literature. *bioRxiv*. <https://doi.org/10.1101/2020.03.20.997783>
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Sanfilippo, P., Hewitt, A. W., & Mackey, D. A. (2018). Plurality in multi-disciplinary research: multiple institutional affiliations are associated with increased citations. *PeerJ*, 6, e5664. <https://doi.org/10.7717/peerj.5664>.
- Saunders, M. E., Duffy, M. A., Heard, S. B., Kosmala, M., Leather, S. R., McGlynn, T. P., et al. (2017). Bringing ecology blogging into the scientific fold: measuring reach and impact of science community blogs. *Royal Society Open Science*, 4(10), 170957. <https://doi.org/10.1098/rsos.170957>.
- Silk, M. J., Harrison, X. A., & Hodgson, D. J. (2020). Perils and pitfalls of mixed-effects regression models in biology. *PeerJ*, 8, e9522. <https://doi.org/10.7717/peerj.9522>.
- Stewart, J. A. (1983). Achievement and Ascriptive Processes in the Recognition of Scientific Articles. *Social Forces*, 62(1), 166–189. <https://doi.org/10.2307/2578354>.
- Tahamtan, I., & Bornmann, L. (2018). Core elements in the process of citing publications: Conceptual overview of the literature. *Journal of Informetrics*, 12(1), 203–216. <https://doi.org/10.1016/j.joi.2018.01.002>.
- Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*, 107(3), 1195–1225. <https://doi.org/10.1007/s11192-016-1889-2>.
- Taylor, D. M. D. (2002). The appropriate use of references in a scientific research paper. *Emergency Medicine*, 14(2), 166–170. <https://doi.org/10.1046/j.1442-2026.2002.00312.x>.
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics work? Twitter and ten other social web services. *PLoS ONE*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>.
- Von Bartheld, C. S., Collin, S. P., & Güntürkün, O. (2009). To each citation, a purpose. *Science*, 323(5910), 36–37. <https://doi.org/10.1126/science.323.5910.36c>.
- Waltman, L., & van Eck, N. J. (2019). Field normalization of scientometric indicators. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 281–300). Cham: Springer. https://doi.org/10.1007/978-3-030-02511-3_11
- Webster, G. D., Jonason, P. K., & Schember, T. O. (2009). Hot topics and popular papers in evolutionary psychology: analyses of title words and citation counts in evolution and human behavior, 1979–2008. *Evolutionary Psychology*, 7(3), 147470490900700300. <https://doi.org/10.1177/147470490900700301>.
- Wright, M., & Armstrong, S. J. (2011). The ombudsman: verification of citations: Faulty towers of knowledge? *Interfaces*, 38, 125–139. <https://doi.org/10.2139/ssrn.1941335>.
- Wu, L. L., Huang, M. H., & Chen, C. Y. (2012). Citation patterns of the pre-web and web-prevalent environments: the moderating effects of domain knowledge. *Journal of the American Society for Information Science and Technology*, 63(11), 2182–2194. <https://doi.org/10.1002/asi.22710>.
- Yu, T., Yu, G., Li, P. Y., & Wang, L. (2014). Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics*, 101(2), 1233–1252. <https://doi.org/10.1007/s11192-014-1279-6>.
- Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, 7(6), 636–645. <https://doi.org/10.1111/2041-210X.12577>.
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2009). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>.
- Zuur, A. F., Savaliev, A. A., & Ieno, E. N. (2012). *Zero inflated models and generalized linear mixed models with R* (2nd ed.). Newburgh: Highland Statistics Limited.